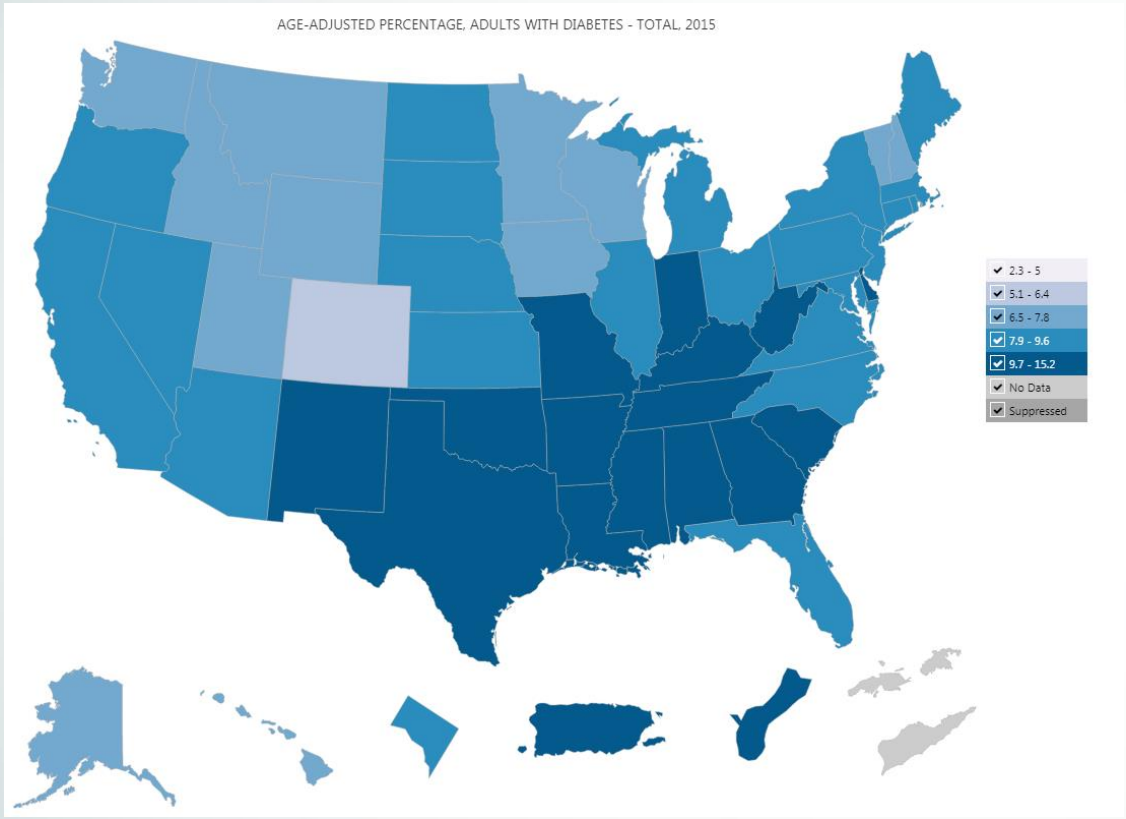
Predicting Unplanned Medical Visits among Patients with Diabetes Using Machine Learning

Arielle Selya, PhD, Sanford Research

Eric L. Johnson, MD, University of North Dakota

SANFORD RESEARCH

Diabetes



- High blood glucose
  - Type 1: no insulin production
  - Type 2: insulin resistance
- 9% of U.S. population
- $245 billion annual costs [1]
  - $176 billion in direct medical costs [1]

Source: https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html

Effects of Diabetes

- Hyper/hypoglycemia [2]
- Vascular complications [2]
  - Cardiovascular disease (CVD)
  - Nerve damage
  - Kidney damage
  - Eye damage
- Infections [3]
  - Soft tissue
  - Respiratory tract
  - Urinary tract

- Patients have many unplanned medical visits [4]

Diabetes and Smoking

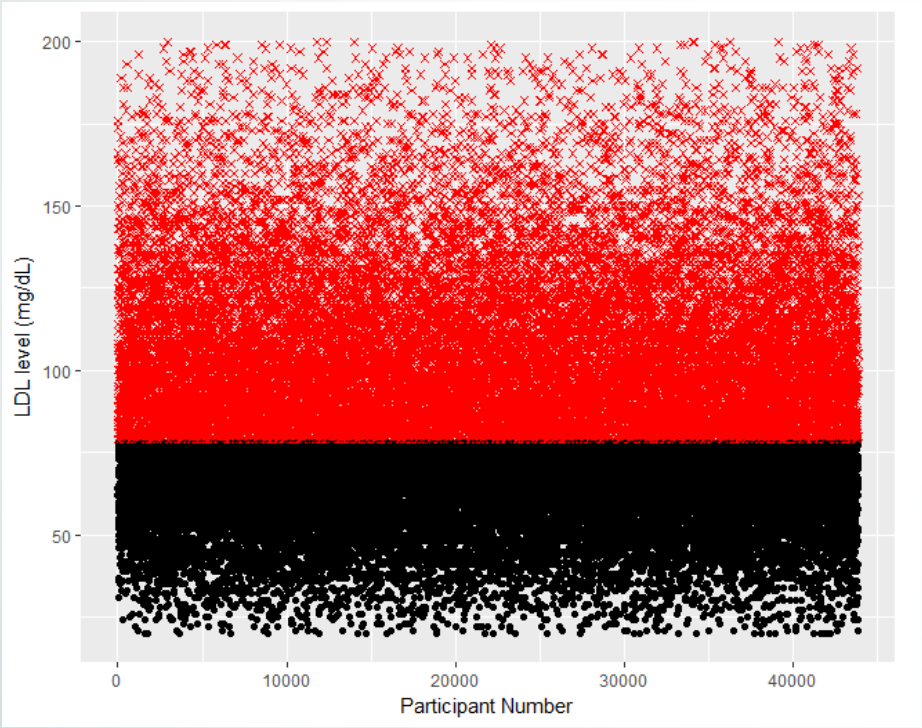
- Smoking exacerbates the complications of diabetes:
  - Decreases glycemic control [1]
  - Increases risk of infection [5]
  - Amplifies CVD risk [6]

Can we predict unplanned visits?

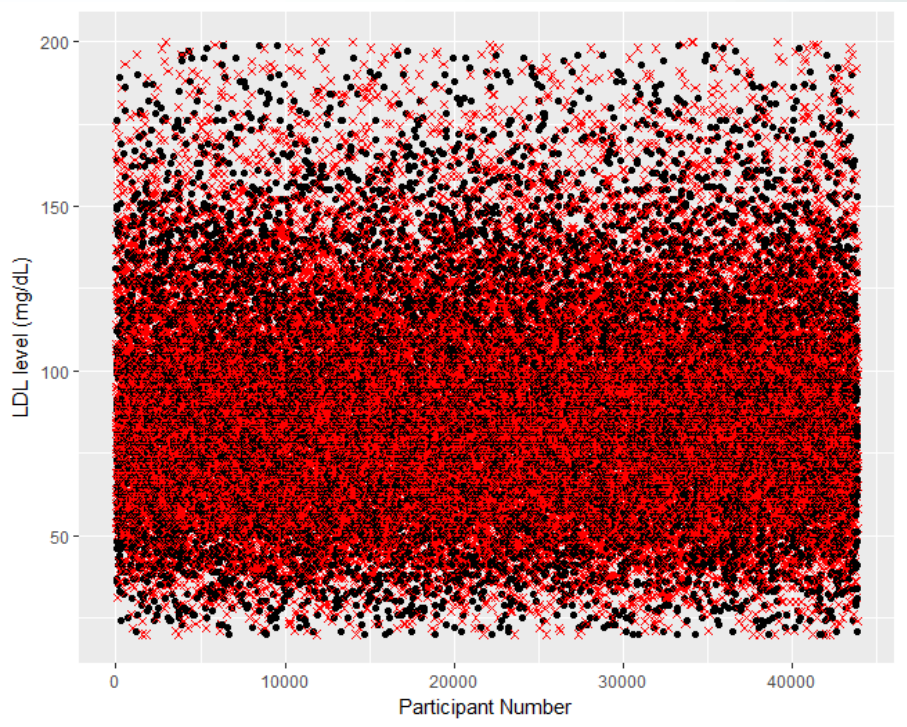
- LACE index for 30-day readmissions

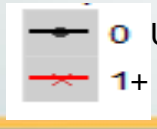
Risk Prediction Challenges

Hypothetical: Perfect prediction



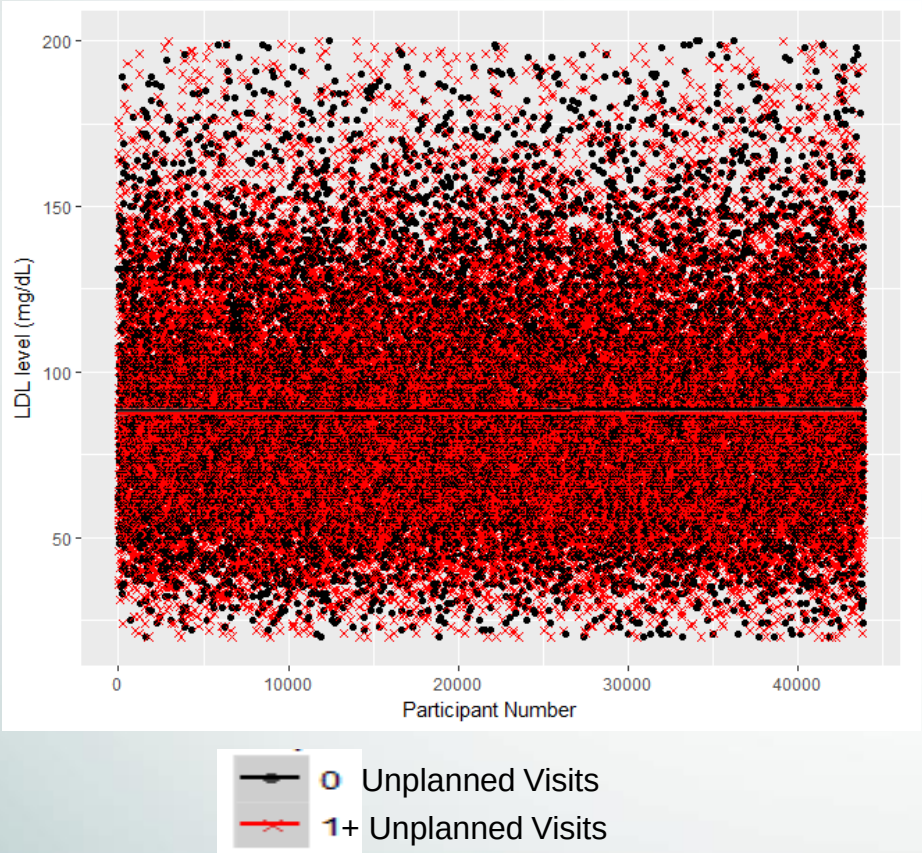
Hypothetical: Random prediction





Risk Prediction Challenges

Actual Data



- Small but statistically significant difference (88.3 vs. 87.5, *p*=.005)
- Statistical differences do not necessarily indicate *predictive ability*!

Machine Learning for Risk Prediction

- Classification task (any vs. no unplanned visits)
  - Linear and quadratic discriminant analysis
  - Support vector machines (SVM)
  - Artificial neural nets (NN)
- Relative to status quo
  - Logistic regression analysis
  - LACE index

Sanford Data Collaborative 2017

- EMR data 2014-16
- N=63,245 patients:
  - Age 18 or over
  - Diabetes diagnosis
  - Zip codes in MN, ND, SD
- Unplanned visits
  - 4 separate types
  - 54.7% had ≥1 unplanned visit

- Predictors:
  - Age
  - Blood pressure
  - Number on “problem list”
  - Number of prescriptions
  - Body mass index (BMI)
  - Cholesterol (HDL, LDL)
  - A1C
  - Ranked smoking status

Unplanned Visits by Smoking Status

	Smokers	Nonsmokers	*p*	Change
Minnesota	58.6%	50.4%	<.0001	+ 8.2%
North Dakota	59.6%	57.5%	.0400	+ 2.1%
South Dakota	59.5%	55.5%	.0003	+ 4.0%

- Patients with diabetes who smoke are more likely to have at least one unplanned visit

Most Common Diagnoses in Unplanned Visits

Smokers with diabetes

Diagnosis	Frequency
R10.xx: Abdominal and pelvic pain	6.6% (*N*=4109)
M54.xx: Dorsalgia	6.1% (*N*=3750)
R07.xx: Pain in throat and chest	3.6% (*N*=2252)
M25.xx: Other joint disorder, not elsewhere classified	3.4% (*N*=2114)
M79.xx: Other and unspecified tissue disorders	3.4% (*N*=2085)
L03.xx: Cellulitis and acute lymphangitis	2.3% (*N*=1453)
E11.xx: Type 2 Diabetes Mellitus	2.1% (*N*=1314)
R05.xx: Cough	2.1% (*N*=1298)
J40.xx: Bronchitis, not specified as acute or chronic	2.0% (*N*=1265)
G43.xx: Migraine	1.9% (*N*=1197)

Nonsmokers with diabetes

Diagnosis	Frequency
R10.xx: Abdominal and pelvic pain	5.4% (*N*=8856)
M54.xx: Dorsalgia	4.3% (*N*=7099)
R07.xx: Pain in throat and chest	4.3% (*N*=6971)
M25.xx: Other joint disorder, not elsewhere classified	3.2% (*N*=5175)
M79.xx: Other and unspecified tissue disorders	3.1% (*N*=5092)
R05.xx: Cough	2.8% (*N*=4547)
L03.xx: Cellulitis and acute lymphangitis	2.2% (*N*=3526)
J40.xx: Bronchitis, not specified as acute or chronic	2.1% (*N*=3422)
J02.xx: Acute pharyngitis	2.0% (*N*=3222)
R51.xx: Headache	1.8% (*N*=3.14)

Evaluating Classifiers

- Cross-validation testing
  - How well did the classifier learn patterns that are truly diagnostic of a category/outcome?
- Confusion matrices

	Predicted class: 0	**Predicted class: 1**
Actual class: 0	Correct rejection	False alarm
Actual class: 1	Miss	Hit

- Average prediction accuracy: mean of correct rejection and hit rates

Risk Prediction Results

Logistic regression

	Prediction: No visits	Prediction: 1+ visit
Actual: No visits	60.5%	39.5%
Actual: 1+ visit	29.8%	70.2%

Average: 65.4%

Best-case classifier (radial-basis support vector machine)

	Prediction: No visits	Prediction: 1+ visit
Actual: No visits	67.8%	32.2%
Actual: 1+ visit	34.1%	65.9%

Average: 66.9% (+1.5 % points)

LACE index for 30-day readmissions:
66.3% hit rate; 53.3% false rejection rate = average 59.8%

Impact of More Accurate Prediction

- For the broader population (not restricted to patients with diabetes):
  - N=379,870 people with 1+ unplanned visit
  - Using SVM over regression correctly identifies N≈3039 people at risk (≈10,000 visits)
- Analyses of cost were not feasible

Clinical Implications

- Can’t conclude causality (from classifiers or regressions)
- Separate treatment from prediction?
- How to extract clinical implications?
  - E.g. what predictor variables, if modified, would lower unplanned visits?
  - Remove patients with certain ranges on modifiable variables, and re-run models

Variables' Impact on Prediction Accuracy

Restricted variable range	New prediction accuracy	Change in accuracy
BMI < 30 (N=15,885)	65.9%	-1.0%
BP < 120/80 (N=11,996)	65.0%	-1.9%
No current smoking (N=38,370)	65.8%	-1.1%
LDL < 130 (N=39,384)	65.8%	-1.1%
HDL > 50 (N=30,058)	65.1%	-1.8%
A1C < 6.5 (N=13,857)	66.2%	-0.7%

High levels of BP and HDL were most informative for predicting unplanned visits

Next Steps

- Data Analysis of Existing Data
  - Adding more variables and refining the model
  - Validate the model with forthcoming data
  - Generate recommendations for clinical targets
- Clinical Research at Sanford
  - Target strong predictors (BP, HDL, smoking) and prospectively look at unplanned visits
  - Identify causal relationships to leverage
  - Automated system for flagging high-risk patients

References

1. American Diabetes Association (2013). Economic costs of diabetes in the U.S. in 2012. *Diabetes Care*;**36**(4):1033-46 doi: 10.2337/dc12-2625.
2. Centers for Disease Control and Prevention (2014). National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States. Atlanta, GA: U.S. Department of Health and Human Services.
3. Standards of Care. American Diabetes Association. Diabetes Care 2016 Jan; 39 (Supplement 1): S1-S2.
4. Washington RE, Andrews RM, Mutter R (2010). Emergency Department Visits for Adults with Diabetes. Statistical Brief #167, Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Agency for Healthcare Research and Quality (US), Rockville (MD).
5. Hofer SE, Miller K, Mermann JM, et al. (2016). International comparison of smoking and metabolic control in patients with Type 1 Diabetes. Diabetes Care, 39(10):e177-e178.
6. McDonald HI, Nitsch D, Millett ERC, Sinclair A, Thomas SL. (2014). New estimates of the burden of acute community-acquired infections among older people with diabetes mellitus: a retrospective cohort study using linked electronic health records. Diabet Med, 31(5):606-614.